

**CSE- 4027 Ex: 8 Assignment****Academic year:** 2020-2021**Semester:** WIN**Faculty Name:** Dr Karthikeyan Saminathan sir**Date:** 10 /11/2021**Student name:** Taran Mamidala**Reg. no.:** 19BCE7346**Working with IRIS Dataset**

1. Read the “sample.txt” text file in R. Print the number of characters, number of digits, number of symbols and number of words in the word file.

```
library(readr)

myData = read_lines("stringdemo.txt", n_max = 1)

charsplit <- strsplit(myData, "")[[1]]

charcount=0

digitcount=0

words = strsplit(myData, " ")[[1]]

for (i in charsplit) {

val=utf8ToInt(i)

if((val>=65 && val <= 90) || (val>=97 && val<=122)){

charcount=charcount+1

}

if(val>=48 && val<=57){

digitcount=digitcount+1

}
```

```
# print(val)

}

symbol=length(charsplit)-(charcount+digitcount)

print(paste("no of chars:",charcount))

print(paste("no of digits:",digitcount))

print(paste("no of symbols:",symbol))

print(paste("no of words:",length(words)))
```

```
> library(readr)
> myData = read_lines("sample.txt", n_max = 1)
> charsplit <- strsplit(myData, "")[[1]]
> charcount=0
> digitcount=0
> words = strsplit(myData, " ")[1]
> for (i in charsplit) {
+   val=utf8ToInt(i)
+   if((val>=65 && val <= 90) || (val>=97 && val<=122)){
+     charcount=charcount+1
+   }
+   if(val>=48 && val<=57){
+     digitcount=digitcount+1
+   }
+ }
> symbol=length(charsplit)-(charcount+digitcount)
> print(paste("no of chars:",charcount))
[1] "no of chars: 5"
> print(paste("no of digits:",digitcount))
[1] "no of digits: 7"
> print(paste("no of symbols:",symbol))
[1] "no of symbols: 19"
> print(paste("no of words:",length(words)))
[1] "no of words: 19"
```

2. Append a new row to the “sample.txt” text file.

```
library(readr)
myData = read_lines("stringdemo.txt", n_max = 1)
write.table(myData, file = "stringdemo.txt", sep = " ",
append = T, quote = F, col.names = F, row.names = F)
```

Alex	25	177	57	F
Lilly	31	163	69	F
Mark	23	190	83	M
Oliver	52	179	75	M
Martha	76	163	70	F
Lucas	49	183	83	M
Caroline	26	164	53	F
Alex	25	177	57	F

### 3. How to read this text file with missing values?

```
myData = read.table(file = "stringdemo.txt", header = TRUE)
print(myData)
print(dim(myData))
print(is.na(myData))
```

```
> myData = read.table(file = "sample.txt", header = TRUE)
> print(myData)
  Alex X25 X177 X57 F
1  Lilly  31  163  69 F
2   Mark  23  190  83 M
3  Oliver  52  179  75 M
4  Martha  76  163  70 F
5   Lucas  49  183  83 M
6 Caroline  26  164  53 F
7   Alex  25  177  57 F
> print(dim(myData))
[1] 7 5
> print(is.na(myData))
  Alex X25 X177 X57 F
[1,] FALSE FALSE FALSE FALSE FALSE
[2,] FALSE FALSE FALSE FALSE FALSE
[3,] FALSE FALSE FALSE FALSE FALSE
[4,] FALSE FALSE FALSE FALSE FALSE
[5,] FALSE FALSE FALSE FALSE FALSE
[6,] FALSE FALSE FALSE FALSE FALSE
[7,] FALSE FALSE FALSE FALSE FALSE
```

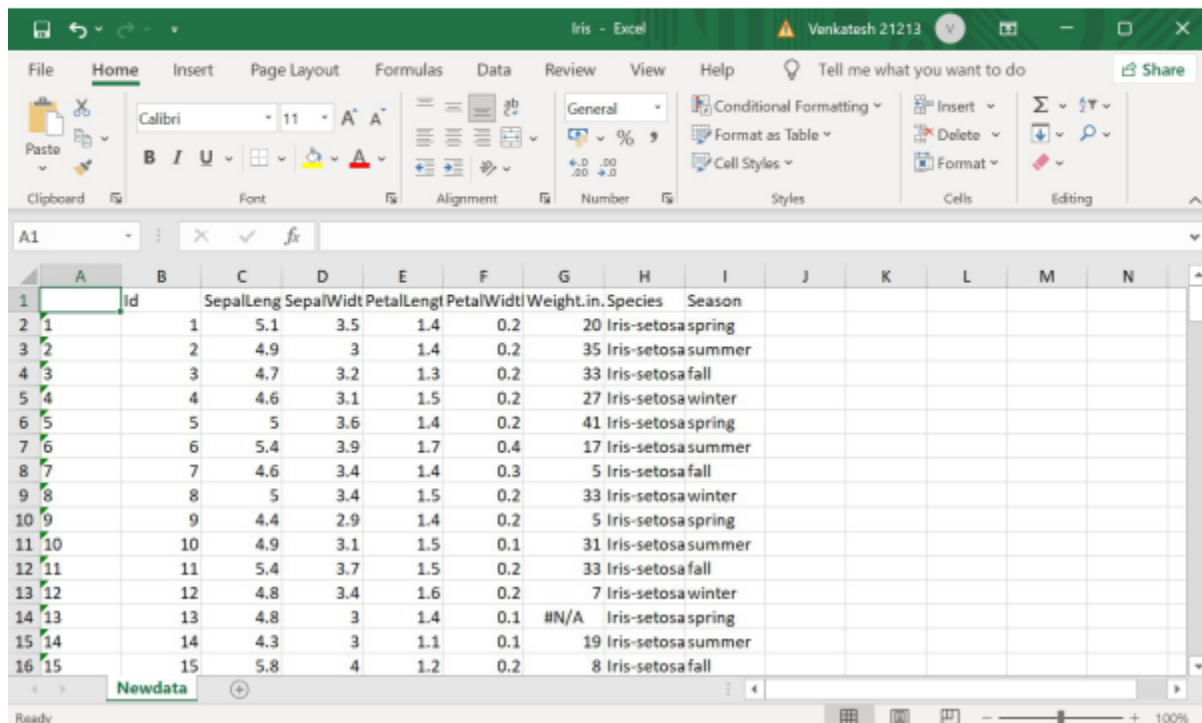
4. Read the Iris dataset from csv file and write into a xlsx file in R.
5. Write the covid data set csv file dataset in the second sheet of the xlsx file created for question2.

```
install.packages("xlsx")
library("xlsx")

dataset = read.csv("Iris.csv")
head(dataset,5)

write.xlsx(dataset,"Iris.xlsx",sheetName = "Newdata")
```

```
> dataset = read.csv("Iris.csv")
> head(dataset,5)
  Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm Weight.in.gm Species Season
1  1           5.1           3.5           1.4           0.2          20 Iris-setosa spring
2  2           4.9           3.0           1.4           0.2          35 Iris-setosa summer
3  3           4.7           3.2           1.3           0.2          33 Iris-setosa fall
4  4           4.6           3.1           1.5           0.2          27 Iris-setosa winter
5  5           5.0           3.6           1.4           0.2          41 Iris-setosa spring
>
```



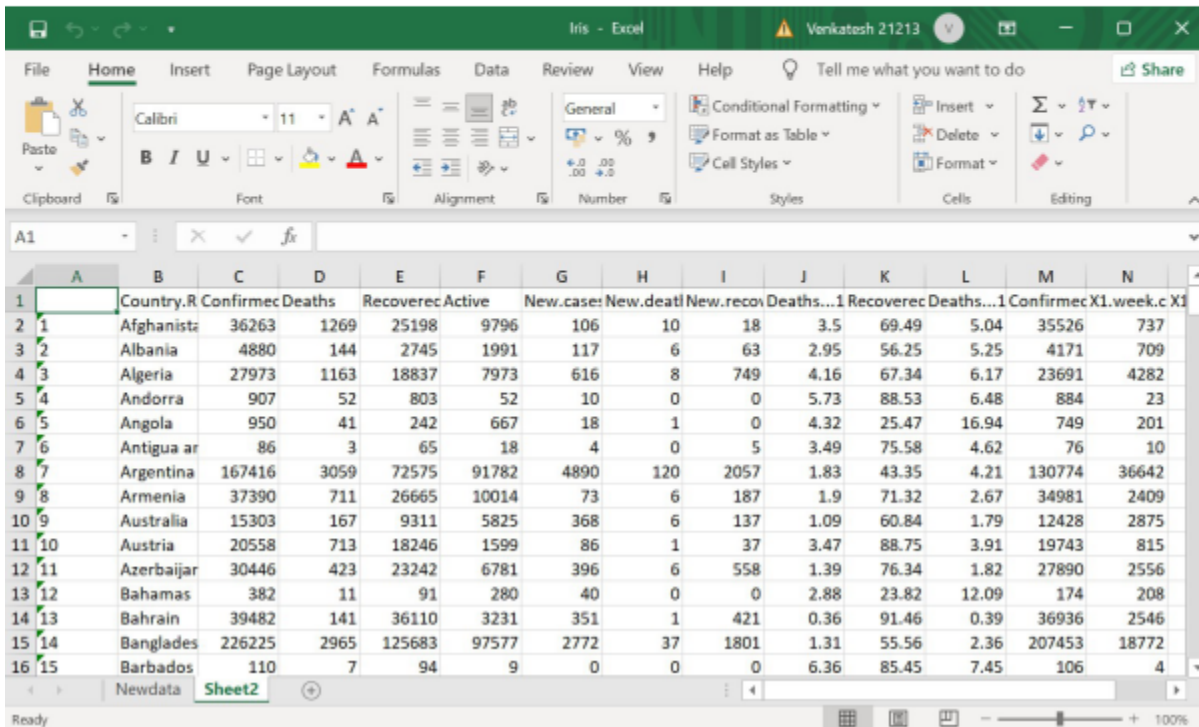
The screenshot shows the Microsoft Excel interface with the 'Iris' workbook open. The 'Newdata' sheet is active, displaying the Iris dataset. The data is organized into columns: Id, SepalLeng, SepalWidt, PetalLengt, PetalWidt, Weight.in, Species, and Season. The first 15 rows of data are visible, corresponding to the first 15 rows of the dataset.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Id	SepalLeng	SepalWidt	PetalLengt	PetalWidt	Weight.in	Species	Season					
2	1	1	5.1	3.5	1.4	0.2	20	Iris-setosa	spring					
3	2	2	4.9	3	1.4	0.2	35	Iris-setosa	summer					
4	3	3	4.7	3.2	1.3	0.2	33	Iris-setosa	fall					
5	4	4	4.6	3.1	1.5	0.2	27	Iris-setosa	winter					
6	5	5	5	3.6	1.4	0.2	41	Iris-setosa	spring					
7	6	6	5.4	3.9	1.7	0.4	17	Iris-setosa	summer					
8	7	7	4.6	3.4	1.4	0.3	5	Iris-setosa	fall					
9	8	8	5	3.4	1.5	0.2	33	Iris-setosa	winter					
10	9	9	4.4	2.9	1.4	0.2	5	Iris-setosa	spring					
11	10	10	4.9	3.1	1.5	0.1	31	Iris-setosa	summer					
12	11	11	5.4	3.7	1.5	0.2	33	Iris-setosa	fall					
13	12	12	4.8	3.4	1.6	0.2	7	Iris-setosa	winter					
14	13	13	4.8	3	1.4	0.1	#N/A	Iris-setosa	spring					
15	14	14	4.3	3	1.1	0.1	19	Iris-setosa	summer					
16	15	15	5.8	4	1.2	0.2	8	Iris-setosa	fall					

6. Differentiate scan() and read\_table using Iris data set.

```
dataset2 = read.csv("COVID_country_wise_latest.csv")
head(dataset2,5)
write.xlsx(dataset2,"Iris.xlsx",sheetName = "Sheet2",append = T)
```

```
> dataset2 = read.csv("COVID_country_wise_latest.csv")
> head(dataset2,5)
  Country.Region Confirmed Deaths Recovered Active New.cases New.deaths New.recovered Deaths...100.Cases Recovered...100.Cases
1  Afghanistan 36263 1269 25198 9796 106 10 18 3.50 69.49
2  Albania 4880 144 2745 1991 117 6 63 2.95 56.25
3  Algeria 27973 1163 18837 7973 616 8 749 4.16 67.34
4  Andorra 907 52 803 52 10 0 0 5.73 88.53
5  Angola 950 41 242 667 18 1 0 4.32 25.47
  Deaths...100.Recovered Confirmed.last.week X1.week.change X1.week...increase WHO.Region
1 5.04 35526 737 2.07 Eastern Mediterranean
2 5.25 4171 709 17.00 Europe
3 6.17 23691 4282 18.07 Africa
4 6.48 884 23 2.60 Europe
5 16.94 749 201 26.84 Africa
```



The screenshot shows the Microsoft Excel interface with the file 'Iris - Excel' open. The 'Sheet2' tab is selected, displaying the COVID dataset. The data is organized into columns: Country.Region, Confirmed, Deaths, Recovered, Active, New.cases, New.deaths, New.recovered, Deaths...100.Cases, Recovered...100.Cases, Confirmed.last.week, X1.week.change, X1.week...increase, and WHO.Region. The first 16 rows of data are visible, starting with Afghanistan and ending with Barbados.

	Country.Region	Confirmed	Deaths	Recovered	Active	New.cases	New.deaths	New.recovered	Deaths...100.Cases	Recovered...100.Cases	Confirmed.last.week	X1.week.change	X1.week...increase	WHO.Region
1	Afghanistan	36263	1269	25198	9796	106	10	18	3.50	69.49	5.04	35526	737	2.07 Eastern Mediterranean
2	Albania	4880	144	2745	1991	117	6	63	2.95	56.25	5.25	4171	709	17.00 Europe
3	Algeria	27973	1163	18837	7973	616	8	749	4.16	67.34	6.17	23691	4282	18.07 Africa
4	Andorra	907	52	803	52	10	0	0	5.73	88.53	6.48	884	23	2.60 Europe
5	Angola	950	41	242	667	18	1	0	4.32	25.47	16.94	749	201	26.84 Africa
6	Antigua and Barbuda	86	3	65	18	4	0	5	3.49	75.58	4.62	76	10	
7	Argentina	167416	3059	72575	91782	4890	120	2057	1.83	43.35	4.21	130774	36642	
8	Armenia	37390	711	26665	10014	73	6	187	1.9	71.32	2.67	34981	2409	
9	Australia	15303	167	9311	5825	368	6	137	1.09	60.84	1.79	12428	2875	
10	Austria	20558	713	18246	1599	86	1	37	3.47	88.75	3.91	19743	815	
11	Azerbaijan	30446	423	23242	6781	396	6	558	1.39	76.34	1.82	27890	2556	
12	Bahamas	382	11	91	280	40	0	0	2.88	23.82	12.09	174	208	
13	Bahrain	39482	141	36110	3231	351	1	421	0.36	91.46	0.39	36936	2546	
14	Bangladesh	226225	2965	125683	97577	2772	37	1801	1.31	55.56	2.36	207453	18772	
15	Barbados	110	7	94	9	0	0	0	6.36	85.45	7.45	106	4	

7. Use iris dataset and plot the normal distribution on all the numerical columns



```
> Iris = scan("Iris.csv", what="character")
Read 153 items
> Iris
[1] "Id,SepalLengthCm,SepalWidthCm,PetalLengthCm,PetalWidthCm,Weight"
[2] "in"
[3] "gm,Species,Season"
[4] "1,5.1,3.5,1.4,0.2,20,Iris-setosa,spring"
[5] "2,4.9,3,1.4,0.2,35,Iris-setosa,summer"
[6] "3,4.7,3.2,1.3,0.2,33,Iris-setosa,fall"
[7] "4,4.6,3.1,1.5,0.2,27,Iris-setosa,winter"
[8] "5,5,3.6,1.4,0.2,41,Iris-setosa,spring"
[9] "6,5.4,3.9,1.7,0.4,17,Iris-setosa,summer"
[10] "7,4.6,3.4,1.4,0.3,5,Iris-setosa,fall"
[11] "8,5,3.4,1.5,0.2,33,Iris-setosa,winter"
[12] "9,4.4,2.9,1.4,0.2,5,Iris-setosa,spring"
[13] "10,4.9,3.1,1.5,0.1,31,Iris-setosa,summer"
[14] "11,5.4,3.7,1.5,0.2,33,Iris-setosa,fall"
[15] "12,4.8,3.4,1.6,0.2,7,Iris-setosa,winter"
```

```
> dataset <- read_table("Iris.csv")

-- Column specification -----
cols(
  `Id,SepalLengthCm,SepalWidthCm,PetalLengthCm,PetalWidthCm,Weight` = col_character()
,
  `in` = col_character(),
  `gm,Species,Season` = col_character()
)

Warning: 150 parsing failures.
row col expected actual file
  1 -- 3 columns 1 columns 'Iris.csv'
  2 -- 3 columns 1 columns 'Iris.csv'
  3 -- 3 columns 1 columns 'Iris.csv'
  4 -- 3 columns 1 columns 'Iris.csv'
  5 -- 3 columns 1 columns 'Iris.csv'
... ..
See problems(...) for more details.
```

```
> dataset
# A tibble: 150 x 3
  Id, SepalLengthCm, SepalWidthCm, PetalLengthCm, PetalWidthCm, Species, Season
  <chr>
1 1, 5.1, 3.5, 1.4, 0.2, 20, Iris-setosa, spring
2 2, 4.9, 3.1, 1.4, 0.2, 35, Iris-setosa, summer
3 3, 4.7, 3.2, 1.3, 0.2, 33, Iris-setosa, fall
4 4, 4.6, 3.1, 1.5, 0.2, 27, Iris-setosa, winter
5 5, 5.3, 3.6, 1.4, 0.2, 41, Iris-setosa, spring
6 6, 5.4, 3.9, 1.7, 0.4, 17, Iris-setosa, summer
7 7, 4.6, 3.4, 1.4, 0.3, 5, Iris-setosa, fall
8 8, 5.3, 3.4, 1.5, 0.2, 33, Iris-setosa, winter
9 9, 4.4, 2.9, 1.4, 0.2, 5, Iris-setosa, spring
10 10, 4.9, 3.1, 1.5, 0.1, 31, Iris-setosa, summer
# ... with 140 more rows
```

a. dnorm()

```
> dnorm(iris$Sepal.Length)
 [1] 8.972435e-07 2.438961e-06 6.369825e-06 1.014085e-05 1.486720e-06
 [6] 1.857362e-07 1.014085e-05 1.486720e-06 2.494247e-05 2.438961e-06
[11] 1.857362e-07 3.961299e-06 3.961299e-06 3.853520e-05 1.977320e-08
[16] 3.513955e-08 1.857362e-07 8.972435e-07 3.513955e-08 8.972435e-07
[21] 1.857362e-07 8.972435e-07 1.014085e-05 8.972435e-07 3.961299e-06
[26] 1.486720e-06 1.486720e-06 5.361035e-07 5.361035e-07 6.369825e-06
[31] 3.961299e-06 1.857362e-07 5.361035e-07 1.076976e-07 2.438961e-06
[36] 1.486720e-06 1.076976e-07 2.438961e-06 2.494247e-05 8.972435e-07
[41] 1.486720e-06 1.598374e-05 2.494247e-05 1.486720e-06 8.972435e-07
[46] 3.961299e-06 8.972435e-07 1.014085e-05 3.171349e-07 1.486720e-06
[51] 9.134720e-12 5.088140e-10 1.830332e-11 1.076976e-07 2.669557e-10
[56] 3.513955e-08 9.601433e-10 2.438961e-06 1.386680e-10 5.361035e-07
[61] 1.486720e-06 1.101576e-08 6.075883e-09 3.317884e-09 6.182621e-08
[66] 7.131328e-11 6.182621e-08 1.977320e-08 1.793784e-09 6.182621e-08
[71] 1.101576e-08 3.317884e-09 9.601433e-10 3.317884e-09 5.088140e-10
[76] 1.386680e-10 3.630962e-11 7.131328e-11 6.075883e-09 3.513955e-08
[81] 1.076976e-07 1.076976e-07 1.977320e-08 6.075883e-09 1.857362e-07
[86] 6.075883e-09 7.131328e-11 9.601433e-10 6.182621e-08 1.076976e-07
[91] 1.076976e-07 3.317884e-09 1.977320e-08 1.486720e-06 6.182621e-08
[96] 3.513955e-08 3.513955e-08 1.793784e-09 8.972435e-07 3.513955e-08
[101] 9.601433e-10 1.977320e-08 4.513544e-12 9.601433e-10 2.669557e-10
[106] 1.144156e-13 2.438961e-06 1.069384e-12 7.131328e-11 2.207990e-12
[111] 2.669557e-10 5.088140e-10 3.630962e-11 3.513955e-08 1.977320e-08
[116] 5.088140e-10 2.669557e-10 5.324148e-14 5.324148e-14 6.075883e-09
[121] 1.830332e-11 6.182621e-08 5.324148e-14 9.601433e-10 7.131328e-11
[126] 2.207990e-12 1.793784e-09 3.317884e-09 5.088140e-10 2.207990e-12
[131] 5.127754e-13 1.118796e-14 5.088140e-10 9.601433e-10 3.317884e-09
[136] 5.324148e-14 9.601433e-10 5.088140e-10 6.075883e-09 1.830332e-11
[141] 7.131328e-11 1.830332e-11 1.977320e-08 3.630962e-11 7.131328e-11
[146] 7.131328e-11 9.601433e-10 2.669557e-10 1.793784e-09 1.101576e-08
> |
```

b. pnorm()



```
> pnorm(iris$Sepal.Width)
 [1] 0.9997674 0.9986501 0.9993129 0.9990324 0.9998409 0.9999519 0.9996631 0.9996631 0.9981342 0.9990324 0.9998922
 [12] 0.9996631 0.9986501 0.9986501 0.9999683 0.9999946 0.9999519 0.9997674 0.9999277 0.9999277 0.9996631 0.9998922
 [23] 0.9998409 0.9995166 0.9996631 0.9986501 0.9996631 0.9997674 0.9996631 0.9993129 0.9990324 0.9996631 0.9999793
 [34] 0.9999867 0.9990324 0.9993129 0.9997674 0.9998409 0.9986501 0.9996631 0.9997674 0.9892759 0.9993129 0.9997674
 [45] 0.9999277 0.9986501 0.9999277 0.9993129 0.9998922 0.9995166 0.9993129 0.9993129 0.9990324 0.9892759 0.9974449
 [56] 0.9974449 0.9995166 0.9918025 0.9981342 0.9965330 0.9772499 0.9986501 0.9860966 0.9981342 0.9981342 0.9990324
 [67] 0.9986501 0.9965330 0.9860966 0.9937903 0.9993129 0.9974449 0.9937903 0.9974449 0.9981342 0.9986501 0.9974449
 [78] 0.9986501 0.9981342 0.9953388 0.9918025 0.9918025 0.9965330 0.9965330 0.9986501 0.9996631 0.9990324 0.9892759
 [89] 0.9986501 0.9937903 0.9953388 0.9986501 0.9953388 0.9892759 0.9965330 0.9986501 0.9981342 0.9981342 0.9937903
 [100] 0.9974449 0.9995166 0.9965330 0.9986501 0.9981342 0.9986501 0.9986501 0.9937903 0.9981342 0.9937903 0.9998409
 [111] 0.9993129 0.9965330 0.9986501 0.9937903 0.9974449 0.9993129 0.9986501 0.9999277 0.9953388 0.9860966 0.9993129
 [122] 0.9974449 0.9974449 0.9965330 0.9995166 0.9993129 0.9974449 0.9986501 0.9974449 0.9986501 0.9974449 0.9999277
 [133] 0.9974449 0.9974449 0.9953388 0.9986501 0.9996631 0.9990324 0.9986501 0.9990324 0.9990324 0.9990324 0.9965330
 [144] 0.9993129 0.9995166 0.9986501 0.9937903 0.9986501 0.9996631 0.9986501
> |
```

c. qnorm()

```
> qnorm(iris$Petal.Width)
 [1] -0.8416212 -0.8416212 -0.8416212 -0.8416212 -0.8416212 -0.2533471 -0.5244005 -0.8416212 -0.8416212 -1.2815516
 [11] -0.8416212 -0.8416212 -1.2815516 -1.2815516 -0.8416212 -0.2533471 -0.2533471 -0.5244005 -0.5244005 -0.5244005
 [21] -0.8416212 -0.2533471 -0.8416212 0.0000000 -0.8416212 -0.8416212 -0.2533471 -0.8416212 -0.8416212 -0.8416212
 [31] -0.8416212 -0.2533471 -1.2815516 -0.8416212 -0.8416212 -0.8416212 -0.8416212 -1.2815516 -0.8416212 -0.8416212
 [41] -0.5244005 -0.5244005 -0.8416212 0.2533471 -0.2533471 -0.5244005 -0.8416212 -0.8416212 -0.8416212 -0.8416212
 [51]      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
 [61]      Inf      Inf      Inf      Inf      Inf      Inf      Inf      Inf      Inf      Inf      Inf
 [71]      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
 [81]      NaN      Inf      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
 [91]      NaN      NaN      NaN      NaN      Inf      NaN      NaN      NaN      NaN      NaN      NaN
 [101]      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
 [111]      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
 [121]      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
 [131]      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
 [141]      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
Warning message:
In qnorm(iris$Petal.Width) : NaNs produced
> |
```

d. rnorm()

```
> rnorm(iris$Petal.Width)
 [1] 0.7090836722 -0.7164011216 0.7185751749 -0.7794460060 -0.2823345024 1.5487647927 0.0623889010
 [8] -0.3955116948 -0.1557497035 -0.5540165787 0.3458317048 -0.3181574869 -1.3649941526 -0.1181016970
 [15] 0.1834243790 1.3684188728 -0.5991164487 -0.3162736847 -0.2817963906 1.7296571673 0.4068058687
 [22] -1.0445311733 -0.4364925552 0.7085399167 -2.2922336517 -0.4925319720 0.0413644134 0.3336465429
 [29] 0.4596835122 2.6928616929 2.5399195743 -0.3155157111 -0.0005452421 0.8810342615 0.0165382732
 [36] 0.2893023642 -0.6273817141 -0.3424542466 0.4016477419 1.0154879344 -0.5084485910 0.7149092069
 [43] -0.5521109135 0.2150021023 -0.6131578789 -0.8531833986 2.2257585798 -1.1310313461 -0.5892043989
 [50] 2.8175013143 -0.3484489383 0.5767482550 -0.7315509534 -0.3870850485 0.0786932272 0.0462138452
 [57] 1.0928200384 0.8473401147 -0.8795869912 -0.4359543590 0.3137235358 0.9981150934 -0.4476541709
 [64] 1.0529304695 1.9226497068 -0.2557432343 -1.1461991676 -0.1626651933 -1.2087412565 -1.3002072823
 [71] 0.7335079709 -0.8409689152 -0.2393117250 -0.3170895263 -1.9637252920 -0.4936328047 0.4476124682
 [78] 0.1391638757 -0.8577151999 -0.6023167726 -0.2859672731 -0.3735876743 -1.5710392907 1.5138543700
 [85] -1.9940315462 -1.5071950329 0.7318796702 0.8796240664 -0.0243705099 -0.3188808057 -0.4898425763
 [92] -1.0080250070 1.1716726876 1.4627323785 0.5276428732 0.1179212212 0.3952572116 0.6305620396
 [99] 1.4681204874 -0.0401068457 0.5880187113 -0.8448422186 -0.1201891630 -0.3114376769 0.4430258657
 [106] -2.2699026543 0.3163582719 -0.3360581320 0.0721116142 0.7526732280 1.0121873384 -0.3823663673
 [113] 0.9578859718 -0.6425700782 -0.6991663439 1.0461148686 0.8561001074 0.9690961799 -0.5389776945
 [120] 0.8499386333 1.4891089198 -0.4252972926 1.8943627215 0.2930361088 0.8981560503 0.3324140503
 [127] -0.5342233419 -0.1950799429 -0.6701928137 1.0144230552 0.5277189327 0.7111642357 0.8263034715
 [134] 0.3673185063 0.3630082315 0.8149989495 0.6811584570 1.2771332677 -0.6794257538 -1.6150712639
 [141] 0.5853788188 -0.6583329084 -0.0471631875 -0.4541953203 0.1657802516 -1.8026854254 -0.7172628982
 [148] -0.3480016541 0.3409382734 -2.0497390852
> |
```